A Computational Strategy for the Solution of Large Linear Inverse Problems in Geophysics

Osni Marques and Tony Drummond
Lawrence Berkeley National Laboratory
Computational Research Division
1 Cyclotron Road, MS 50F-1650
Berkeley, CA 94720-8139
oamarques@lbl.gov, ladrummond@lbl.gov

Don Vasco
Lawrence Berkeley National Laboratory
Earth Sciences Division
1 Cyclotron Road, MS 90R1116
Berkeley, CA 94720-8139
dwvasco@lbl.gov

Abstract

This paper discusses the use of singular values and singular vectors in the solution of large inverse problems that arise in the study of physical models for the internal structure of the Earth. In this study, the Earth is discretized into layers and the layers into cells, and travel times of sound waves generated by earthquakes are used to construct the corresponding physical models. The underlying numerical models lead to sparse matrices with dimensions up to $1.3x10^6$-by-$3x10^5$. Singular values and singular vectors of these matrices are then computed and used in the solution of the associated inverse problems and also to estimate uncertainties. The paper outlines the formulation adopted to model the Earth and the strategy employed to compute singular values and singular vectors, shows results for two models that have been studied, comments on the main computation issues related to the solution of these problems on high performance parallel computers, and discusses future improvements of the adopted computational strategy.

1. Introduction

The solution of many practical problems requires the computation of eigenvalues λ and eigenvectors *z* of an *n*-by-*n* sparse matrix *A*, that is to say, the computation of non-trivial solutions of the problem

$$Az = \lambda z. \qquad (1)$$

Without being exhaustive, eigenpairs are used to study configuration of molecules, electric circuits, neutron fluxes in nuclear power plants, dynamic properties of structural models, and also for extraction of features in biometric applications. Our problem corresponds to a variation of (1). In particular, we are interested in the singular value decomposition problem (SVD) [1] of an *m*-by-*n* matrix *G*,

$$G = U\Sigma V^T, \qquad (2)$$

where $U = [u_1\ u_2 \ldots u_r]$ and $V = [v_1\ v_2 \ldots v_r]$ are *m*-by-*r* and *n*-by-*r* orthogonal matrices, respectively, and $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots \sigma_r),\ \sigma_1 \geq \sigma_2 \geq \ldots \sigma_r,\ r = \min(m,n)$. The columns of *U* are the *left singular vectors*, the columns of *V* are the *right singular vectors*, and the entries in Σ are the *singular values*. The SVD allows for the computation of generalized (or pseudo) inverses, which are important for the solution of inverse problems [2]. The generalized inverse of *G* is defined as [3]

$$G^\dagger = V\Sigma^{-1}U^T. \qquad (3)$$

The solution of inverse problems is required in applications such as satellite navigation, medical tomography, image enhancement and seismic tomography. In such applications, given a set of physical data, the goal is to estimate a set of model parameters describing the problem at hand. Also important is to determine limits of the inferences that can be drawn from the data. In practical analyses, *G* and its generalized inverse are approximated by a truncated or low-rank SVD, that is to say,

$$G \approx G_k = U_k\Sigma_k V_k^T, \qquad (4a)$$

$$G^\dagger \approx G_k^\dagger = V_k\Sigma_k^{-1}U_k^T, \qquad (4b)$$

where $U_k = [u_1\ u_2 \ldots u_k],\ V_k = [v_1\ v_2 \ldots v_k]$, and $\Sigma_k = \mathrm{diag}(\sigma_1, \sigma_2, \ldots \sigma_k)$, $k < r$.

Problems (1) and (2) are strongly related to each other because *u*, *v* and σ can be obtained by means of the eigenvalue problems [1]

$$(G^TG)v = \lambda v, \qquad (5a)$$

$$(GG^T)u = \lambda u, \qquad (5b)$$

where $\lambda = \sigma^2$. One of these two eigenvalue problems is of order *m*, the other is of order *n*; whichever problem is of larger order has *m*-*n* additional eigenvalues equal to zero. Alternatively, one can compute the SVD by solving the augmented eigenvalue problem [1]

$$\begin{bmatrix} 0 & G \\ G^T & 0 \end{bmatrix} \begin{Bmatrix} u \\ v \end{Bmatrix} = \lambda \begin{Bmatrix} u \\ v \end{Bmatrix}, \qquad (6)$$

where $\lambda = \pm\sigma$. This problem also has *m*-*n* additional eigenvalues equal to zero.

Similarly to *A* in problem (1), *G* is usually sparse and its dimensions can be very large. In addition, several hundred singular values and singular vectors may be required in some applications. Therefore, preserving and exploiting sparsity is critical to finding solutions efficiently, which means that G^TG and GG^T in (5) should not be explicitly computed since they are likely to be full.

In the last few decades, models of the 3-D seismic structure of the Earth have become increasingly sophisticated [4]-[6]. It is known that the 3-D distribution of density, temperature and the *P* (compressional) and *S* (shear) velocities within the Earth have a direct relationship to its dynamics. A great variety of studies have been made possible due to a continuous accumulation of seismic data around the globe, advancements both in instrumentation and data analysis, our ability to model processes in the Earth, such as mantle and core convection[a], and the development and implementation of efficient algorithms for dealing with inverse problems [7]. Since there is no direct means of sampling the interior of the Earth, information must be derived indirectly, mainly from measuring the travel time of seismic waves as they cross through the Earth. The signal source is an earthquake and the receivers are more than 1,000 seismographic stations around the globe.

This paper is concerned with some of the calculations that are required to obtain a physical model for the internal structure of the Earth. In this study, the Earth is discretized into layers and the layers into cells. The velocity of wave propagation in each cell is represented by a set of parameters that account for anisotropy, location and correction of sources and receivers, etc. These parameters are written in matrix form and the goal is to fit the model by means of some known data. Travel times of sound waves generated by earthquakes are used to construct the model. The underlying numerical models lead to sparse matrices with up to $1.3\text{x}10^6$ rows (the number of travel times used), $3\text{x}10^5$ (the number of parameters in the model) and $5\text{x}10^7$ non-zero entries.

[a] Basically, the Earth is formed by a crust, a mantle and a core. The crust is about 8 km thick under the oceans and 40 km under the continents. Below the crust and roughly 3000 km thick is the mantle, which is separated into upper and lower portions. Below the mantle are Earth's outer and inner cores.

Several numerical strategies have been proposed to assess the reliability of physical models of the Earth. Vasco *et al.* [5], for example, use a formulation that requires an LU decomposition [1] of a full matrix for the calculation of resolution and covariances, and which can therefore be prohibitive for large models because of storage and computer time requirements. Alternatively, variations of the least-squares QR decomposition (LSQR) [1] for the calculation of resolution and covariances have been proposed by Yao *et al.* [13] and Zhang and McMechan [13]. Also important in our application is the determination of spatial resolution and uncertainties, but these may not be very well determined by LSQR type methods. In the present work we focus on the calculation and use of singular values and singular vectors to solve inverse problems and to estimate uncertainties. Although our matrices are sparse, such problems are difficult to deal with since a large subset of singular values and singular vectors may be required for a proper estimation of spatial resolution and uncertainties.

In the following sections we outline the formulation adopted to image the Earth and the strategy employed to compute singular values and singular vectors, give results for two models that have been studied, comment on the main computation issues related to the solution of these problems on high performance parallel computers, and discuss future improvements of the adopted computational strategy.

2. Methodology

In this section we outline the representation of Earth's structure, the estimation of velocity variations in the entire Earth and the assessment of the resulting estimates. The procedure is akin to constructing a CAT scan in medical imaging.

2.1. Representation of the structure of the Earth

Several classes of parameters have to be taken into account in order to represent the complexities of wave propagation through the Earth [4]-[6]. Let us consider a travel time for a phase of type *k* (*P*,*S*,*PP*,*PKPbc*…) observed at station *i*, associated with earthquake *j*, δt_{ijk}. In our model we solve for shifts in the depth, latitude and longitude of the earthquake and shifts in the earthquake origin time, δh_j^l, $l = 1,2,3,4,$ and include station delays for compressional and shear phases read at each station, δs_j^l, $l = 1,2$. The velocity is assumed to be constant in each cell and the unknowns represent deviations from the average layer velocity, δv. We model the velocity of sound waves traveling in the Earth.

2.2. Robust estimation

The relationship between the change in arrival time and the structural parameters of the Earth is non linear. However, it is thought that the deviations from a purely radial variation are small. Therefore, a Taylor series expansion truncated to the first order produces a linear equation relating arrival time deviations to perturbations in Earth's structure [6][7],

$$\delta t_{ijk} = \sum_{l=1}^{n_{3D}} \frac{\partial t_{ijk}}{\partial v^l}\delta \mathrm{v}^l_k + \sum_{l=1}^{4} \frac{\partial t_{ijk}}{\partial v^l}\delta h^l_j + \sum_{l=1}^{n_s} \frac{\partial t_{ijk}}{\partial s^l}\delta s^l_{ik} + \sum_{l=1}^{n_b} \frac{\partial t_{ijk}}{\partial b^l}\delta b^l + \sum_{l=1}^{n_a} \frac{\partial t_{ijk}}{\partial a^l}\delta a^l, \quad (7)$$

for the *k*-th phase arrival time associated with event *j*, recorded at station *i*, where n_s, n_b and n_a are functions of the ray path. Given a large collection of arrival times from various source and receiver combinations, we can infer variations in the structure of the Earth. The relationship between the vector of traveltime deviations, δt, the vector of velocity perturbations, δv, hypocentral location changes, δh, station corrections, δs, boundary deflections, δb, and variations in inner core anisotropy, δa, may be written in matrix form as

$$\delta t = \mathrm{V}\delta v + \mathrm{H}\delta h + \mathrm{S}\delta s + \mathrm{B}\delta b + \mathrm{A}\delta a, \quad (8)$$

where V, H, S, B and A contain the partial derivatives of the travel time with respect to velocity variation, earthquake locations, station corrections, boundary deflections and inner core anisotropy deviations, respectively [6][7].

A least-squares technique can be used to solve equation (8) [1][7]. However, the excessive number of outliers often found in travel time data sets may lead to a violation of the basic assumptions underlying the least squares (normally distributed errors). The approach we adopt is based upon the minimization of the l^p norm of the residual vector [8]. In our particular application, we set $p = 1.25$, which is derived from an analysis of our composite data set [9]. Defining $\mathrm{M} = [\mathrm{V}\ \mathrm{H}\ \mathrm{S}\ \mathrm{B}\ \mathrm{A}]$ and $\delta x = [\delta v\ \delta h\ \delta s\ \delta b\ \delta a]^T$ we write the l^p residual norm minimization problem in a compact form as

$$\min \sum_{i=1}^{m}\sum_{j=1}^{n} \left| \mathrm{m}_{ij}\delta x_j - \delta t_i \right|^p, \quad (9)$$

where *m* is the total number of constraints and *n* is the total number of parameters. The solution of (9) can be carried out by repeatedly solving a weighted system of linear equations [11]

$$\mathrm{N}^{1/2}\mathrm{M}\delta x = \mathrm{N}^{1/2}\delta t, \quad (10)$$

where the weights or scaling contained in $\mathrm{N}^{1/2}$ are varied systematically. The matrix $\mathrm{N}^{1/2}$ is diagonal and its *i*-th element is

$$\mathrm{n}_i^{1/2} = |\sum_{j=1}^{n} \mathrm{m}_{ij}\delta x_j - \delta t_i|^{p-2}. \quad (11)$$

The solution of (10), δx, is used to compute the set of scaling factors in (11), the equations are rescaled and (10) is solved again. The procedure is repeated until satisfactory convergence is achieved. If we define $G = \mathrm{N}^{1/2}\mathrm{M}$ and $\delta d = \mathrm{N}^{1/2}\delta t$, we can write,

$$G\delta x = \delta d, \quad (12)$$

whose approximate solution, $\delta\tilde{x}$, can be computed by means of the generalized inverse of *G*,

$$\delta\tilde{x} = V_k\Sigma_k^{-1}U_k^T\delta d. \quad (13)$$

2.3. Model Assessment

Model parameter resolution is a quantitative measure of averaging or blurring inherent to model parameter estimates. If we assume that a "true" solution, $\delta\hat{x}$, exists for (12), then we can substitute δd in (13) to obtain

$$\delta\tilde{x} = G_k^{\dagger}G_k\delta\hat{x} = V_kV_k^T\delta\hat{x} = R\delta\hat{x}. \quad (14)$$

The elements of a row of *R*, the *resolution matrix*, are averaging coefficients that quantify the contribution that all parameters make to the estimates. For a rank-*k* approximation of *G*, the averaging results from retaining only *k* singular values and singular vectors, rather than the full spectrum (*r* singular values and singular vectors).

The model covariance matrix contains measures of the uncertainties associated with the model parameters as well as the mapping of the uncertainty between parameters. In particular, the diagonal elements are the variances corresponding to the estimates. Assuming that the data are uncorrelated with uniform variance σ_d^2, we can write the model parameter covariance matrix as

$$C_m = \sigma_d^2 G_k^{\dagger}(G_k^{\dagger})^T = \sigma_d^2 V_k\Sigma_k^{-2}V_k^T. \quad (15)$$

If one is only interested in calculating elements of *R* and C_m, the amount of computation can be reduced because only a subset of the singular values and corresponding right singular values is required for those calculations. Also, to find a solution for (13) one could alternatively use a least squares algorithm [9][11]-[13].

3. The Computation of a Partial SVD

In this section we outline the approach used to compute *k* singular values and singular vectors of *G*. Algorithms for computing the SVD can be downloaded from Netlib [15], for instance. However, the algorithms available in Netlib are intended for the computation of either a small subset of singular values and singular vectors of sparse matrices [16] or the complete set of singular values and singular vectors of full matrices [17][18].

The method we use in our applications is based on a block version of the Lanczos algorithm [18] implemented by one of us. Given a symmetric *n*-by-*n* matrix *A* and $Q = [q_1\ q_2 \ldots q_p]$, where Q is orthogonal and *p* is the *block size*, the block Lanczos algorithm generates a basis for the *Krylov subspace* defined as $\mathcal{K}(A,Q,j) = span(Q, AQ, \ldots A^{j-1}Q)$. In this subspace, $A^{j-1}Q$ converges towards the eigenvectors associated with the *p* dominant eigenvalues of *A* as *j* increases. One of the advantages of generating a basis for $\mathcal{K}(A,Q,j)$ is that more than *p* approximations for eigensolutions of *A* can be obtained as *j* increases. Another important feature of the algorithm is that *A* is accessed only as a matrix-vector product for the basis generation process.

For the computation of the SVD, one can set $A = G^T G$ or $A = GG^T$ as indicated in (5). Also, one can choose the problem of smaller order to compute the shorter set of singular vectors, say (5a) to compute V if $m > n$ (which is the case in our applications). Given V, U can be obtained as $U = AV\Sigma^{-1}$. The algorithm we use is summarized below as Algorithm 1 [20]. In that algorithm, the product AQ_j is carried out in two steps, *a* and *b*. Steps *c* to *e* correspond to Gram-Schmidt orthogonalizations [1], and in step *f* a modified Gram-Schmidt orthogonalization procedure can be used to factor *R* as $Q_{j+1}\mathrm{B}_{j+1}$, where Q_{j+1} is orthogonal and B_{j+1} is an upper triangular matrix [19]. Then, after *j* iterations the vectors Q can be arranged as $\mathrm{Q}_j = [Q_1\ Q_2 \ldots Q_j]$, where Q_j is orthogonal. Also, $\mathrm{Q}_j^T A \mathrm{Q}_j = T_j$, where T_j is a block symmetric tridiagonal matrix,

$$T_j = \mathrm{tridiag}\begin{pmatrix} & \mathrm{B}_2^T & & \cdots & \mathrm{B}_j^T & \\ \mathrm{A}_1 & & \mathrm{A}_2 & \cdots & & \mathrm{A}_j \\ & \mathrm{B}_2 & & \cdots & \mathrm{B}_j & \end{pmatrix}, \qquad (16)$$

which means that the projection of problem (1), or equivalently (5a), into Q_j is given by T_j. An approximate solution $(\tilde{\lambda}, \tilde{z})$ for (1), or equivalently

Algorithm 1. Block Lanczos Algorithm.

Set $Q_0 = 0, B_1 = 0$. Choose Q_1, $Q_1^T Q_1 = I$.

for $j = 1,2,\ldots j_{\max}$

a) $S \leftarrow GQ_j$

b) $R \leftarrow G^T S$

c) $R \leftarrow R - Q_{j-1}\mathrm{B}_j^T$

d) $\mathrm{A}_j \leftarrow Q_j^T R$

e) $R \leftarrow R - Q_j \mathrm{A}_j$

f) $R \rightarrow Q_{j+1}\mathrm{B}_{j+1}$

end

$(\tilde{\lambda}, \tilde{v})$ for (5a), is given by the *Ritz value* $\tilde{\lambda} = \theta_i$ and by the *Ritz vector* $\tilde{z} = \mathrm{Q}_j s_i$, where (θ_i, s_i) is the solution of the eigenvalue problem $T_j s_i = \theta_i s_i$, $i = 1,2,\ldots\ jxp$. Usually, the extreme eigenvalues (and corresponding eigenvectors) of T_j lead to good approximations of extreme eigenvalues (and corresponding eigenvectors) of the original *n*-by-*n* problem with *j*<< *n*. Convergence can be monitored through the residual norm $\eta = \| A\tilde{z} - \tilde{\lambda}\tilde{z} \|$, which can be approximated at very low cost by means of the quantities B_{j+1} and s_i computed as the algorithm progresses [19].

In finite precision arithmetic, a loss of orthogonality among the vectors of the basis generated by the Lanczos algorithm is generally observed after some steps, which leads to the convergence of spurious eigenvalues and eigenvectors. However, several strategies can be implemented to monitor and keep the orthogonality within a certain level, such as selective orthogonalization and partial orthogonalization [19]. The implementation of such strategies implies in additional steps in Algorithm 1 (see [20] for details).

The main disadvantage of Algorithm 1 is that by using $A = G^T G$ the condition number of *G* is squared and this can lead to numerical difficulties if small singular values are wanted. However, this may be advantageous when only the largest singular values are of interest, and this is what is sought in our applications. Another possibility for the computation of the SVD of *G* is the bidiagonal formulation of the Lanczos algorithm applied to (6) [1]. In this case, however, convergence of small singular values is slow because of the location of those values in the center of the spectrum $\lambda = \pm\sigma$. In addition, vectors of length *m*+*n* are required in problem (6) and this

can be a hurdle for very large problems. Alternatively, the bidiagonal Lanczos algorithm can be rearranged to generate two bases of vectors of length *m* and *n* [1].

4. Implementation Details

In our applications we use a parallel, MPI based version, of Algorithm 1. A sparse format storage scheme is used to store *G*: one (integer) array stores the number of nonzero entries in each row, one (integer) array stores the column indices, and one (real) array stores the corresponding entries. In our implementation one processor reads *G* and distributes its rows among the participating processors. If needed, another redistribution is performed to guarantee that each processor holds roughly the same number of entries of *G*; therefore, the number of rows of *G* per processor may differ. This (re)distribution is important to assure a good load balance in the parallel matrix-vectors products involving *G*. On the other hand, the matrix Q_j generated by Algorithm 1 is full and the memory it requires is proportional to $j_{\max}$. In our implementation Q_j is distributed also by rows among the available processors. In this case, however, each processor stores *m*/*npe* rows of Q_j, where *m* is the number of rows and *npe* is the number of processors (a similar strategy is used for *V*, which is also full). Because of the storage scheme adopted for Q_j, we need to group the rows of *G* contained in each processor in blocks of *m*/*npe* columns each. This facilitates the computation of *S* and *R* in Algorithm 1 but requires a prior (quick) sorting and reordering of the arrays that store the column indices and entries of *G*. With this strategy we also greatly reduce cache misses in the (sparse) matrix products involving *G*.

In order to compute a large number of singular values and singular vectors for large models we use Algorithm 1 only to generate Q. Tests for convergence are not performed, which means that we solve $Ts = \theta s$ and compute singular vectors in a post-processing phase. This strategy allows us to save computing time and the memory required to store the eigenvectors of *T*, as well as the Ritz vectors $\tilde{z} = Qs$. In addition, we can save Q and restart the algorithm to add more vectors to it if needed, therefore obtaining more approximations for singular values and vectors (and at the same time introducing a checkpoint capability). We use parallel IO features implemented in MPI-2 to read and write all relevant information for restarting into a (binary) file.

For the parallel solution of the (block) tridiagonal eigenvalue problem, we use ScaLAPACK, which provides algorithms for dense linear algebra calculations and which is one of the tools available in the US Department of Energy's (DOE) Advanced Computational Software (ACTS) Collection [21]. The ACTS Collection is a set of computational tools developed primarily at DOE laboratories and is aimed at simplifying the solution of common and important computational problems. The use of ACTS tools reduces the development time for new codes and the tools provide functionality that might not otherwise be available. As a matter of fact, we foresee the use of other ACTS tools in the applications discussed in this work, in particular Global Arrays for performing data distribution, and TAU for a comprehensive performance analysis and tuning of the most time consuming parts of our codes (see next section). For more information on these tools we refer the reader to [21].

5. Numerical Results

In this section we show numerical results related to two models of the Earth. Table 1 shows the dimensions of the numerical models and the number of non-zero entries in the corresponding matrices. In the first model, the Earth is parameterized by 22 depth layers of 1136 cells each. The lateral dimensions of the cells are 6^o-by-6^o at the Equator. In the second model, the Earth is parameterized by 12 layers of 1136x4, 3^o-by-3^o, cells each in the mantle and 10 layers of 1136, 6^o-by-6^o, cells each in the core.

All computations were performed in single precision on a 696 processors CRAY T3E-900, 256 Mbytes of memory on each processor, available at the DOE's National Energy Research Scientific Computing Center (NERSC). Table 2 shows the number of steps, the total wall-clock time (in seconds) and the time breakdown (in %) for the computation of 500 singular values and singular vectors of model 1 using Algorithm 1, for five values of *p*. We used 32 processors and $\eta < 10^{-6}$. The matrix-vector products dominate the costs, followed by the computation of converged Ritz vectors, solution of the reduced eigenvalue problems, and reorthogonalization operations. Reading and distributing *G* among the processors takes roughly 2 minutes. We would like to point out that similar calculations with *p*=1 required several weeks on a Sun workstation [7]. Therefore, the above timings indicate a speed-up of orders of magnitude.

Using the strategy described in Section 4 we were able to compute up to 20000 vectors for model 2 by doing four restarts, using from 64 processors (10000 vectors, 6 hours of wall-clock time) to 256 processors (2000 vectors, 2 hours of wall-clock time). That number of vectors requires roughly 50 Gbytes of disk space. Figure 1 shows the distribution of $\sqrt{\tilde{\lambda}_i}$ for model 2, for four values of *k*. The right end of the curves become flat as *k* increases, which means that small singular values begin to converge. In general, the first third part of each curve, i.e., $(\theta_i, s_i), 1 \le i \le k/3$, lead to good approximations of

Table 1. Characteristics of two models examined.

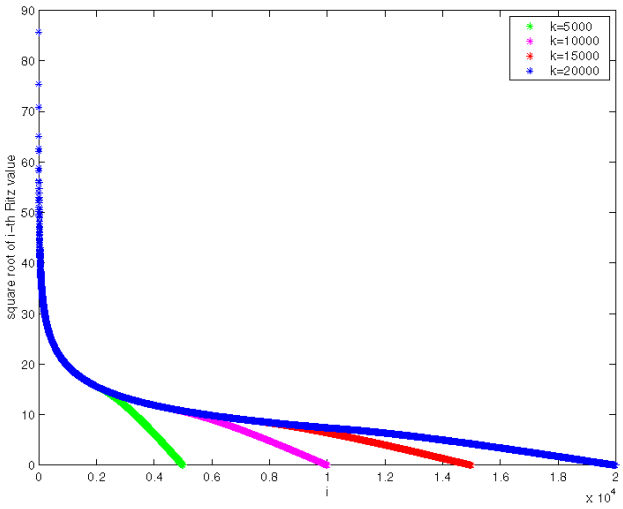
model	*m*	*n*	*non zeros*
1	846968	96300	28587210
2	1433102	307134	47968477

singular values and singular vectors, the remaining s_i lead to linear combinations of singular vectors that are orthogonal to the ones that have converged. We have explored this property in our calculations, by using $\hat{V}_k = \mathrm{Q}_k S_k$, $S_k = [s_1\ s_2 \ldots s_{kxp}]$, for the computation of R, rather than V_k.

In Figure 2 we plot $\mathrm{diag}(R)$ for model 2. As k increases R becomes more diagonal dominant, indicating where the model has been well resolved. We chose a cut-off of 1/1000 of the peak spectral amplitude in the calculations, which means that only the vectors corresponding to the values above the cut-off were used to compute R.

The model parameter resolution, $\mathrm{diag}(R)$, for portions of the mantle for *P* waves with k=9965 is shown in Figure 3, darker tones signify higher resolution. In general, Lanczos resolution is greater in the major subduction zones encircling the Pacific and beneath the continents of the Northern Hemisphere. The co-location

Figure 1. Ritz values distribution for model 2, *k*=5000, 10000, 15000 and 20000.



of seismic sources and seismographic stations in the Pacific subduction zones and tectonically active continental regions is primarily responsible for the well-resolved circum-Pacific velocity heterogeneity. Lack of a significant number of stations in the world's ocean basis results in generally poor Lanczos resolution beneath the Pacific, Atlantic and Indian Ocean basis. The velocity heterogeneity we estimate is shown in Figure 4 also for *P*

Figure 2. Diagonal entries of *R* for model 2, *k*=4986, 9965, 14935 and 19890.

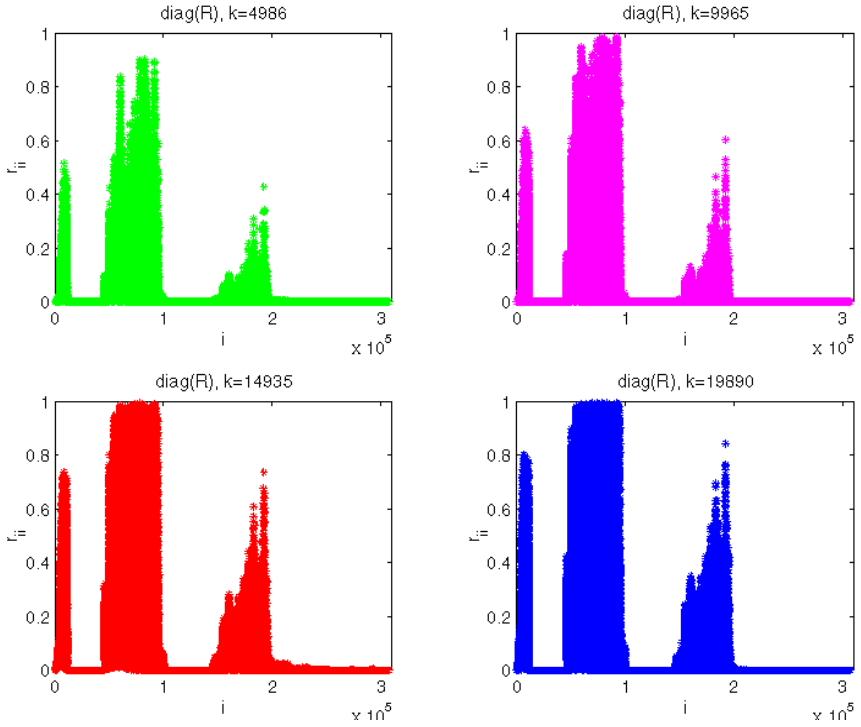


Table 2. Time breakdown for the computation of 500 singular values and singular vectors.

iterations and timings	*p=1*	*p=2*	*p=3*	*p=4*	*p=5*
iterations	1293	671	460	352	289
total time	1270	1190	1110	1100	1170
matrix-vector products (AQ_j)	49.6%	50.5%	53.5%	54.3%	52.6%
basis generation (steps *c* to *f*)	0.1%	0.2%	0.3%	0.4%	0.8%
reorthogonalization[19,20]	12.2%	11.5%	8.1%	7.7%	7.1%
reduced problem ($Ts = \theta s$)	19.7%	16.4%	15.0%	13.1%	13.2%
Ritz vectors (Qs)	18.3%	21.4%	23.0%	24.5%	26.3%

waves. The shallow mantle (35-200 km) is dominated by high velocities associated with cratons and a circum-Pacific low-velocity ring. For a detailed discussion on these results we refer the reader to [8] and [10].

6. Conclusions

With the advent of a large set of seismic data collected in the last decades, as well as the implementation of efficient numerical algorithms on high performance parallel computers, very large inverse problems in Geophysics can now be tackled. Although we showed results obtained on CRAY T3E-900, our codes have been ported to an IBM SP and we are currently in the process of porting them to a Linux cluster. This will facilitate comparisons, exchange of models, and the use of our codes by researchers that have access to more modest computational resources.

In this paper we showed a few examples where a SVD plays an important role in the proper estimation of spatial resolution and uncertainties. It remains to be investigated, however, how our calculations compare, in terms of computational costs and accuracy, with other approaches which use least squares based techniques for the computation of resolutions and covariances such as those proposed in [11]-[13].

7. Acknowledgements

This work was supported by a Laboratory Directed Research and Development Grant, US Department of Energy, under Contract No. DE-AC03-76SF00098. The information presented here does not necessarily reflect the position or the policy of the Government and no official endorsement should be inferred.

All computations were performed at the NERSC facilities (*http://www.nersc.gov*), Lawrence Berkeley National Laboratory (*http://www.lbl.gov*).

8. References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 3rd edition, 1996.

[2] R. Parker, *Geophysical Inverse Theory*, Princeton University Press, 1994.

[3] S. L. Campbell and C. D. Meyer Jr., *Generalized Inverses of Linear Transformations*, Dover Publications, 1991.

[4] J. Pulliam, D. W. Vasco and L. R. Johnson, *Tomographic inversions for mantle P wave velocity structure based on the minimization of l^2 and l^1 norms of international seismological centre travel time residuals*", J. of Geophysical Research, 98, pp. 699-734, 1993.

[5] D. W. Vasco, J. Pulliam, and L. R. Johnson, *Formal inversion of ISC arrival times for mantle P-velocity structure*, Geophysical J. Int., 113, pp. 586-606, 1993.

[6] D. W. Vasco, L. R. Johnson and J. Pulliam, *Lateral variations in mantle velocity structure and discontinuities determined from P, PP, S, SS, and SS-S_dS travel time residuals*, J. of Geophysical Research, 100, pp. 24037-24059, 1995.

[7] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Inc., 1984.

[8] D. W. Vasco, L. R. Johnson and O. A. Marques, *Global Earth Structure: inference and assessment*, Geophysical J. Int., 137, pp. 381-407, 1999.

[9] G. Nolet, *Seismic wave propagation and seismic tomography*, in Seismic Tomography, pp. 1-23, ed. Nolet, G., and Reidel, D., Norwell, MA, 1987.

[10] D. W. Vasco and L. R Johnson, *Whole Earth structure estimated from seismic arrival times*, J. of Geophysical Research, 103, pp. 2633-2671, 1999.

[11] J. A. Scales, A. Gersztenkorn and S. Treitel, *Fast l_p resolution of large, sparse, linear systems, application to seismic travel time tomography*, J. Comp. Physics, 75, pp. 314-333, 1988.

[12] G. Nolet, R. Montelli and J. Virieux, *Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems*, Geophysical J. Int., 138, pp. 36-44, 1999.

[13] Z. S. Yao, R. G. Roberts and A. Tryggvason, *Calculating resolution and covariance matrices for seismic tomography with the LSQR method*, Geophysical J. Int., pp. 886-894, 138, 1999.

[14] J. Zhang and G. A. McMechan, *Estimation of resolution and covariance for large matrix inversions*, Geophysical J. Int., pp. 409-426, 121, 1995.
[15] Netlib, *http://www.netlib.org*.
[16] R. Lehoucq, D. Sorensen and C. Yang, *ARPACK User's Guide: Solution of Large Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods*, SIAM, 1998.
[17] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney and D Sorensen, *LAPACK User's Guide*, 3rd edition SIAM, 1999.
[18] S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker and R. C. Whaley, *ScaLAPACK User's Guide*, SIAM, 1997.
[19] B. N Parlett, "*The Symmetric Eigenvalue Problem*", SIAM, 1998.
[20] O. A. Marques, *BLZPACK User's Guide*, *http://www.nersc.gov/~osni*.
[21] The DOE ACTS Collection, *http://acts.nersc.gov*.
[22] B. L. N. Kennett, E. R. Engdahl and R. Buland, *Constraints on seismic velocities in the Earth from traveltimes*, Geophysical J. Int., 122, pp. 108-124, 1995.

Figure 3. Resolution estimates for the mantle, P waves.

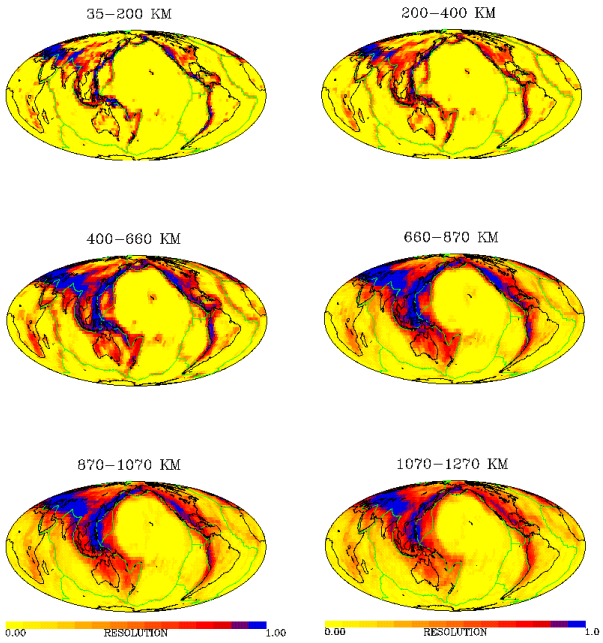


Figure 4. Compressional velocity estimates for the mantle. The lateral variation in velocity are displayed in per cent deviation from the background *ak135* velocity model [22].